

# Data Intake Report

Name: Cab Industry

Report date: 13/08/2022

Internship Batch: LISUM12

Data intake: Mohamad Eyad Abras

Storage Location: Github

## Tabular data details:

<b>Total number of observations</b>	<359392>
<b>Total number of files</b>	<1 , Cab_Data >
<b>Total number of features</b>	<7>
<b>Base format of the file</b>	<.csv >
<b>Size of the data</b>	<21.5 mb>

<b>Total number of observations</b>	<43171>
<b>Total number of files</b>	<1 , Customer_ID>
<b>Total number of features</b>	<4>
<b>Base format of the file</b>	<.csv >
<b>Size of the data</b>	<1 mb>

<b>Total number of observations</b>	<20>
<b>Total number of files</b>	<1 , City>
<b>Total number of features</b>	<3>
<b>Base format of the file</b>	<.csv >
<b>Size of the data</b>	<561 bytes>

<b>Total number of observations</b>	<440098>
<b>Total number of files</b>	<1 , Transaction_ID>
<b>Total number of features</b>	<3>
<b>Base format of the file</b>	<.csv >
<b>Size of the data</b>	<8.58 mb>

**Proposed Approach:**

**Customer\_ID column is used in both files** ( Transaction\_ID and Customer ID) and that helps with merging these 2 files together , the Transaction\_id column can be found in the Cab\_Data file as well , finally, we can merge the cities info from the City.csv file since we have the column City in both City.csv and Cab\_Data.csv.

And now we have our final data set that contains the above mentioned 4 files in one single file.